Chapter 9

Predicting the Peacetime Performance of Military Officers: Officer Selection in the Papua New Guinea Defence Force

Louis P. Bau & Murray J. Dyck

This article reports an evaluation of the officer selection procedures of the Papua New Guinea Defence Force based on the performance of 195 officers. Results indicate that a linear combination of selection variables, including psychometric measures, academic achievement, interview impressions, and hinds-on performance tests, was not predictably related to the criterion measure of officer performance. Post-hoc validation checks suggest that only one of the psychometric tests used in selection meets basic construct validation criteria and that the criterion measure of officer performance is unreliable. No evidence was found to support the predictive validity of selection procedures.

The history of systematic officer selection in the Papua New Guinea Defence Force (PNGDF) goes back to 1957 when a feasibility study into the development of a psychometric-based selection system was conducted by the Australian Army, of which the Pacific Islands Regiment, the precursor to the PNGDF, was then a part (McElwain & Griffiths, 1957). The study resulted in the development of test specifications and the appointment of personnel to develop the selection system. By 1965, a number of tests had been modified for use in Papua New Guinea (PNG) and a set of selection procedures, also including demographic variables, hands-on performance tests, and interview impressions, was implemented (Ord, 1966). Later, just prior to Papua New Guinean independence in 1975, the tests that been used by the Australian Army were replaced by locally developed psychological tests (Hicks, 1973a; Ord, 1957, 1959, 1967a, 1967b, 1971b; Preston, St George, & St George, 1974a). Since independence, selection procedures have remained unchanged.

Prior to its inception and for several years afterwards, considerable effort was invested to "validate" elements of the officer selection system (Ord, 1957, 1959, 1968). In particular, individual ability tests (also used by other PNG institutions) were assessed against training and nonmilitary performance criteria (cf. Ord, 1967a). On the other hand, no attempt was made to evaluate specific predictor variables, or selection procedures as a whole, against the criterion they were meant to predict: performance as an officer. Rather, it was assumed that a valid measure of cognitive ability, or a valid measure of clerical speed and accuracy, would predict officer performance (McElwain, 1967). Whether that assumption is ever warranted remains a moot point (cf. Hicks, 1981; Hutton, 1981; Jackson & Watangia, 1980; St George & Preston, 1980, 1981a, 1981b); the purpose of this study is to evaluate whether it was warranted in the case of the PNGDF.

The PNGDF officer selection system

Officer selection in the PNGDF proceeds by a process of elimination; candidates who fail to achieve a designated standard or who do not possess some requisite attribute are rejected. In brief, candidates must meet minimum physical, health, age, sex (male), and education standards, must not possess a criminal record, must be single, and must have "good" character references (usually headmasters' reports). Beyond these minimum requirements, candidates must achieve satisfactory results on a series of psychological tests, give satisfactory responses during a structured interview, and achieve a passing grade from the "Force Officer Selection Board" (based on observation of candidates' performance on "real-life" officer tasks).

Psychological tests

The test battery includes two measures of verbal ability, one measure of numerical ability, one measure of reasoning ability, and a measure of clerical speed and accuracy. For each test, minimum performance expectations have been established; each candidate must meet the minimum standard on each test,

Verbal and numerical ability

The Pacific Reading Comprehension Test (RC) was developed by the Psychological Services Branch of PNG from the Australian Council for Educational Research (ACER) Test of Reading Attainment (Ord, 1967b). A study of grade 10 school leavers found that RC scores were moderately correlated with mid-year and final examination results and with the Pacific Word Knowledge Test the same study found the RC to have KR20 reliability of .84, and four month test-retest reliability of .78 (Psychological Services Branch, 1982; see also Price, 1973). By contrast, two studies of teaching college students found RC scores to be unrelated to grades in English, Education, Teaching of English as a Second Language, Math, and Science courses (Preston, St George, & St George, 1974b). In a study assessing the validity of RC for selecting administrative college students, RC scores were significantly correlated with "Public Service Higher Certificate", results, and with first term grades in, two of three diploma programs (Psychological Services Branch, no date). However, the text notes that correlations were corrected for restriction of range and coarse grouping without indicating how these "corrections" were made (Psychological Services Branch, no date; St George & Preston, 1980).

The RC's developer appears to have relied heavily on the test's derivation from the ACER test, "whose reliability and validity as measures of English attainment are already well established" (Ord, 1967b, p. 13). Similarly, other investigators have sanguinely reported that "no item analyses were carried out" and "no attempt was made to gather estimates of school performance as one approach to concurrent validation" (Hicks, 1973b, p. 3).

The Pacific Word Knowledge Test (WK) was developed by the Public Service Commission of PNG from the ACER Word Knowledge Test (Ord, 1967b). In the two studies referred to above, WK was found to have KR20 reliability of .86 and four month test-retest reliability of .82 (Psychological Services Branch, 1982); WK was moderately correlated with Public Service Higher Certificate results and was significantly correlated with first term grades in one of three administrative college programs (Psychological Services Branch, no date; see also Hicks & Bowlay, 1974; St George & Preston, 1980). In two studies of teaching college students, WK was significantly correlated with English and Education grades, but unrelated to other course grades in one study, and not significantly correlated with any grades in the second study (Preston, St George, & St George, 1974c). Preston et al. (1974c) also report that WK was moderately correlated with other tests of word knowledge, but unrelated to measures of reasoning ability.

The Numerical Ability Test (NA) was adapted by the PNGDF from the Australian Army Psychology Corps Arithmetic Achievement Test and assesses knowledge of arithmetic, algebra, and geometry; arithmetic (NAI) and mathematics (NA2) subscales are scored separately. We have been unable to locate a manual for this test and no other information on its validity or reliability is available (see also Price, 1984, and St George & Preston, 1980).

Reasoning ability/general intelligence

The Pacific Reasoning Series Test (RA) was developed by the Public Service Commission of PNG. Regarded as a measure of general intelligence (Hicks, 1969), it is based on the Australian Army Psychology Corps Reasoning Test; the purpose of the adaptation was to reduce western cultural content and to use instructions deemed more appropriate to the PNG context. In a study of preliminary year university students, RA was significantly correlated with two other tests of reasoning ability, but not with tests of numerical ability, reasoning number series, matrix completion, and 17 other abilities measures; neither was RA significantly correlated with performance on mid-term examinations in English, Mathematics, Science, and History (Ord, 1971a; see also Preston, St George,

& St George, 1974d). In a separate study, RA was moderately correlated with Public Service Higher Certificate results and with first term grades in two of three administrative college programs (Psychological Services Branch, no date). The reliability of the test according to the KR20 formula has been found to range from .85 (Ord, 1971a) to .92 (Preston et al., 1974d). Split half reliabilities of .84 and .92 (Spearman-Brown corrected) have been reported, as has a 15 month (average) test-retest reliability of .73 (Preston et al., 1974d).

Clerical speed and accuracy

The Speed and Accuracy Test (SA) is an adaptation by the PNGDF of the Australian Army Psychology Corps Speed and Accuracy Test. Subtests are based on separate scores for number (SAI) and name (SA2) checking. No test manual is available and no other information on the validity or reliability of this test has been located.

Structured interview

The purpose of interviewing candidates is to obtain both demographic information and other data which may suggest leadership qualities on the one hand, and anti-authoritarian or anti-social attitudes and history of social instability on the other. No checks on the reliability or validity of either the interview procedure used or the information collected have been conducted (cf. Mumford & Owens, 1987; Schmitt & Robertson, 1990; Wiesner & Cronshaw, 1988).

Force officer selection board

The selection board procedure is derived from one used by the Australian Army Psychology Corps. The procedure involves observing and evaluating candidates' performance in a variety of 11 officer-relevant" activities (e.g., survival planning following hypothetical shipwreck, and escaping from "enemy fire" by crossing, with equipment, an actual crocodile-infested and swift-flowing river) and by conducting a panel interview with each candidate. Board members are experienced officers. No checks on the reliability or validity of board ratings have been conducted (cf. Sackett, 1987; Klimoski & Brickner, 1987).

What constitutes a good military officer?

The point of an officer selection system is to identify those candidates who are likely to succeed in the role of a military officer. But what constitutes success as a military officer, or, more prosaically, just what is it that the predictor variables are meant to predict?

Although the primary function of a military is to fight and win wars, the capacity of an officer to conduct war is not systematically evaluated in the PNGDF. Rather, all officers are evaluated on an annual basis by their immediate supervisor by means of an Officer Evaluation Report (OER). The OER requires ratings of an officer in six multi item categories: personal qualities, proficiency in administrative skills, proficiency in management of personnel, proficiency in planning and directing, proficiency in primary appointment (occupation), and general ability. OER ratings form the basis for decisions about promotion and further training opportunities; ratings are the primary means by which the PNGDF defines its "good" officers.

Like most of the selection variables, the OER was introduced to the PNGDF by the Australian Army and has not undergone substantial change since its introduction. Neither the validity nor the reliability of the OER ratings has been evaluated (cf. DeNisis & Williams, 1988; Sulsky & Balzer, 1988).

Evaluating the PNGDF officer selection system

Ideally, the way to evaluate any set of selection procedures would be to compare individuals selected by the procedures with individuals rejected by the procedures on the performance of the tasks for which individuals are selected. This ideal method was not available to us and so we opted to evaluate relationships between selection variables and performance variables in the population of officers selected by existing procedures.

There are several shortcomings in such a method. The most important shortcoming is that the absence of a comparison group (the unselected) precludes certainty about whether or not selection procedures are useful. For example, if no relationship is observed between a selection variable and a performance variable, it is arguable that the range restrictions inherent in selection procedures are a sufficient explanation of the lack of relationship (cf. Alexander, 1988; Gross & Fleishman, 1987; Gross & McGanney, 1987). It could even be argued that the lack of a relationship would support the validity of selection procedures insofar as a function of the procedures is to exclude unwanted sources of performance variance. But within the terms of this argument, it would be impossible to make any negative conclusion about the validity of the procedures; both finding and not finding the "expected" relationship would validate the procedures.

Although we recognize that "certainty" is limited by our method, we do not accept that it precludes drawing valid inferences. Selection procedures derive from the belief that there exists a relationship between a selection variable and a performance variable. These presupposed relationships between predictor and criterion variables have never been evaluated in the case of the PNGDF officer selection system, and so the appropriate null hypothesis is that no relationship between the respective variables exists.

Method

Subjects

Between 1974 and 1988, 435 candidates were recruited as officers in the PNGDF. We attempted to obtain selection and performance data for all of these officers, but this was not possible. Information on 30 currently serving senior officers was withheld from us, as was information on a further 58 officers who had been discharged for disciplinary reasons. Finally, the files of some 152 officers were found to be substantially incomplete and were dropped from the study. These exclusions resulted in usable data being available for 195 officers, or 44% of the population of PNGDF officers selected by the system.

Predictor variables

In addition to the psychological tests described above, another seven variables were available for study. These included years of formal education (EDUC), school achievement scores in final year tests of English (ENG), mathematics (MATH), and science (SQ, Force Officer Selection Board rating (FOSB), Psychologist's rating (PSYCH), and Officer Intelligence Rating (01R).

The variable OIR is a derived score. The officer intelligence rating is a categorization of RA scores based on decile ranks in a standardization sample. The variable PSYCH reflects the psychologist's subjective impression (rating) as a member of the Force Officer Selection Board and is thus not independent of FOSB ratings generally.

Outcome variables

As a performance criterion, we opted to use the measure that is used by the PNGDF, namely, the Officer Evaluation Report. Although there are many good reasons for believing that the OER is not a good measure of officer performance (cf. Borman, 1983; Landy & Farr, 1980), the fact that it is the PNGDF's operational definition of a "Good Officer" makes it an unassailable practical criterion.

For the purpose of this study, we utilized two OERs for each officer: the OER for the subject's first year of service as an officer (OERI) and the OER for the subject's final, or most recent, year of service as an officer (OER2). The use of OERI means that the performance of all subjects following a fixed period of service was available for study; the use of OER2 means that a measure of ultimate performance was available. These decisions meant that the initial OERs Were completed between 1975 and 1987; current or most recent OERs were completed between 1976 and 1988.

As already noted, an OER consists of ratings within each of six categories. Items within each category were summed to yield a category score; category scores were summed to yield a total score.

Thus, for each officer there were potentially 14 outcome variables, seven from his first and seven from his final year of military service.

Procedure and data analysis

After obtaining the necessary permissions from the PNGDF, personnel files were reviewed and the relevant information extracted and recoded where necessary. Data analyses were performed with the BMDP statistical software program (Dixon, 1985). Nominal alpha was set at .05 for multivariate analyses incorporating a control for experimental error rate, and at .01 otherwise.

Results

Our question was: Are predictor variables positively, substantially, and significantly related to the performance measures? Our initial approach to the data was to perform multiple regression analyses that incorporated all of the predictor variables and each of the major outcome variables (first and most recent OER total scores). For the first OER, the results indicated no significant or substantial relation to the linear combination of predictor variables (ROER I = .29, p > .05). For the second OER, a significant relationship was observed between the performance measure and a linear combination of selector variables (ROER2 = .37, p < .05). However, the negative weightings of several of the regression coefficients (e.g., RC, SA2, RA, 01R, EDUC, SCI, and FOSB) suggest that the result is not so straightforward as it might first appear.

A review of the univariate correlation matrix of all variables shows that several individual predictors - including reasoning ability, officer intelligence rating, level of education, and force officer selection board rating - are significantly related to the most recent OER (but, as with the multiple regression, are unrelated to the first OER; see Table 1). However, the sign of these "significant" predictors indicates that the correlations between predictor and performance variables are the opposite of those expected. Although officers are chosen on the basis of their relatively high intelligence, reasoning ability, level of education, and so forth, among those selected to be officers, it is the relatively unintelligent, uneducated, and unreasoning person who is most highly valued by performance raters.

Finally, in the case of most predictor variables, no relationship with performance variables was observed; indeed, no selection variable was significantly related to the selected officers' first evaluation reports (see Table 1). Reading comprehension, word knowledge, numerical ability, psychologist's recommendation, English, math, and science competence, inter alia, are all unrelated to officer performance. Thus, the variables used to select officers by the PNGDF are either unrelated to officer performance or are related in a manner contrary to selection standards. How can this be?

Table 1.
Product-Moment Correlations between Selection Variables and First and Most-Recent Officer
Evaluation Reports

	OER1	OER2
Word Knowledge	-05	-10
Reading Comprehension	09	-04
Arithmetical Ability	-02	-01
Mathematical Ability	03	-01
Reasoning Ability	-11	-20*
Number Checking	02	-11
Name Checking	00	-00
Level of Education	-10	-26*
English Achievement	-15	-01
Science Achievement	-04	-03
Math Achievement	-09	-01
Force Officer Selection	-11	-20*
Board Ratings Psychologist's Ratings	-02	-09

*significant at .0 1 level, two-tailed, df > 186

Post-hoc validation checks

Abilities measures

The selection measures are clearly invalid vis-a-vis the OER criterion; are these measures also lacking in construct validity independent of their selection function? In order to assess this possibility, further analyses were done. Specifically, where possible the convergent and discriminate validity of a measure was assessed against other concurrent or criterion variables RC (reading comprehension), WK (word knowledge), ENG (English achievement), NA1 (arithmetical ability), NA2 (mathematical ability), MATH (math achievement), SC1 (science achievement), RA (reasoning ability), SA1 (number checking), SA2 (name checking), EDUC (level of education). The first validation check entailed assessing the specific ability measures against each other, and against levels of education, reasoning ability, officer intelligence rating, and school results. The relevant matrix is reported in Table 2 and the results suggest that few of the ability measures are valid.

For the Pacific Word Knowledge Test (WK), Pacific Reading Comprehension Test (RC), and Numerical Ability Test (NA1, NA2), all measures for which a concurrent and a criterion variable were available, there was no evidence of a convergent discriminate pattern. Word knowledge and reading comprehension were both unrelated to school English language scores (ENG); numerical ability was unrelated to mathematics achievement (MATH). Thus, there is no evidence to support the validity of these three tests.

The Pacific Reasoning Series Test (RA) fares rather better. Of the tests under review here, RA was the only measure to correlate positively and, in all cases but two, significantly with all abilities measures, with level of education, and with school test results. In other words, RA behaves in the manner expected of a measure of general intelligence. Thus, despite the absence of other measures of general intelligence, present results support the construct validity of RA.

Table 2

Product-Moment Achievement, and Correlations between Ability Level of Education

	RC	WK	ENG	NAI	NA2	MATH	SCI	RA	SAI	SA2
WK	08	1.0								
ENG	-10	04	1.0							
NA1	-24*	25*	06	1.0						
NA2	13	07	08	52*	1.0					
MATH	-10	-03	73*	02	06	1.0				
SCI	-12	-04	72*	04	03	82*	1.0			
RA	03	36*	18	27*	30	24*	21*	1.0		
SA1	-13	24*	08	30*	08	05	08	24*	1.0	
SA2	-14	21*	07	29*	08	03	06	23*	97*	1.0
EDUC	-10	41*	-11	32*	19*	-12	-15	39*	20*	22*

*significant at .0 1 level, two-tailed, N = 195

Finally, there were insufficient data to properly evaluate the Speed and Accuracy Test (no concurrent or criterion variables). However, the significant correlations between the two forms of this test and both reasoning ability and level of education (and with WK and NAI) are consistent with the presumed measurement aims of the test.

In summary, the post-hoc validation checks of the abilities measures provide relatively unambiguous support for the construct validity of only one of the psychometric selection variables: reasoning ability (RA).

Outcome measures
Finding little evidence to support the construct validity of selection variables may itself explain why officer performance was unpredictable. On the other hand, the measures of officer performance may also account for some of the lack of relationship.

Because the OER is the sole means by which the PNGDF evaluates its officers, there are no other readily accessible criteria against which the validity of the OER can be directly assessed. But because the reliability of any measure constitutes a limit to its validity, we were able to assess the theoretical limit to OER validity by assessing the reliability of OER ratings. Test-retest reliability was determined by correlating OERI with OER2. The result, a product-moment coefficient of $r = .17$, $p > .01$, suggests a rather low upper limit to the predictive power of any selection variable, no matter its construct validity or other characteristics. Insofar as officer performance as measured by the OER is highly unstable, it is then inherently unpredictable.

Discussion
There is no evidence that officer selection procedures in the PNGDF are valid; indeed, there is evidence that they may be counterproductive. Granted that OER ratings are the standard of officer performance, selection procedures might well be altered to select relatively unintelligent and uneducated candidates for officer training. However, because the OER standard of officer performance is itself unreliable, and hence invalid, it would be a pointless exercise to work towards identifying variables that predict OER ratings.

Although there are many grounds on which this study can be criticized, including the lack of a control group, our definition of school examination results as criterion variables, the inconsistent interval in our analyses of OER test-retest reliability, inter alia - what is remarkable to us is the consistency with which the results fail to support the hypotheses on which selection procedures are based. In other words, the results of this study are consistent with the null hypothesis of no relationship, whether the relationship in question is between selection and performance variables, selection and concurrent variables, and selection and other construct-related criterion variables. Furthermore, in those few cases where a relationship between selection and performance variables was observed, the direction of the correlation was opposite to that predicated by selectors.

Of course, one can speculate that there are factors extraneous to the selection system that partially account for our failure to support selection hypotheses. For example, in a developing country like PNG where skilled workers are in short supply, there are strong incentives for the "best" officers to leave the defence force in favour of the civilian sector. Thus, it could be argued that only the relative dullards would remain within the system. But even this kind of post-hoc explanation is unsatisfactory. Negative relationships between intelligence (RA) and performance (OER2) could, after all, only be observed among those in the service; even if the better educated officers subsequently left the military, it was their final year's OER result that formed the basis for the observed negative correlation.

The reasons why the PNGDF officer selection procedures are invalid may be debated, but what is certain is that if a selection system is to be useful in the future, it must be a different system to that used at present. A good place to begin the development of a new system would be to develop new criteria for evaluating officer performance. No set of selection variables can be expected to provide accurate predictions of performance if performance measures are unreliable.

References

Alexander, R. (1988). Group homogeneity, range restriction, and range enhancement effects on correlations. *Personnel Psychology, 41*, 773-778.

Borman, W. (1983). Implications of personality theory and research for the rating of work performance in organizations. In F. Landy, S. Zedeck, & J. Cleveland (Eds.), *Performance measurement and theory* (pp. 127-165). Hillsdale NJ: Erlbaum Associates.

DeNisi, A., & Williams, K. (1988). Cognitive approaches to performance appraisal. In K Rowland & G. Ferris (Eds.), *Research in personnel and human resource management* (pp. 109-156). Greenwich CT: JAI Press.

Dixon, W. (Ed.) (1985). *BMDP statistical software*. Berkeley: University of California Press.

Eaton, N., Hanser, L., Shields, J. (1987). Validating selection tests against job performance. In J. Zeidner (Ed.), *Human productivity enhancement* (pp. 382-438). New York: Pragger.

Gross, A., & Fleishman, L. (1987). The correction for restriction of range and nonlinear regressions: An analytic study. *Applied Psychological Measurement, 11*, 211-217.

Gross, A., & McGanney, M. (1987). The restriction of range problem and nonignorable selection processes. *Journal of Applied Psychology, 72*, 604-610.

Hicks, R. (1969, December). *The validity of psychological test programmes in Papua New Guinea.* A paper presented at Aptitude Testing-Appraisals Stage 2, Australian Institute of Management, Port Moresby, PNG.

Hicks, R. (1973a). Assessing verbal attainments in Papua New Guinea, 1966-1972. *New Guinea Psychologist, 5*, 98-105.

Hicks, R. (1973b, November). *The 1972 norming and test development project: A report.* Port Moresby: Psychological Services Branch, Public Service Board, Papua New Guinea. (National Archives & Public Records Services of Papua New Guinea, Series 1049, Box. 4252, File no. 269)

Hicks, R. (1981). Comments on St George & Preston's "Psychological testing in Papua New Guinea: A critical appraisal of the work of the Psychological Services Branch." *Australian Psychologist, 16*, 275-277.

Hicks, R., & Bowlay, D. (1974). Some comparative validation figures in predicting initial success in technical studies: A note. *New Guinea Psychologist, 6*, 110-113.

Hutton, M. (1981). Letter to the editor. *Australian Psychologist, 16*, 281.

Jackson, P., & Watangia, K. (1980). Reply to: Psychological testing in Papua New Guinea: A critical appraisal of the work of the Psychological Services Branch. *Australian Psychologist*, *15*, 501-503.

Klimoski, R., & Brickner, M. (1987). Why do assessment centers work? The puzzle of assessment center validity. *Personnel Psychology, 40*, 243-260.

Landy, F., & Farr, J. (1980). Performance ratings. *Psychological Bulletin*, *87*, 72-107.

McElwain, D. (1967, August). *Cognitive factors in counselling and selection.* Paper presented at the Technical Meeting on the Selection and Assessment of Pupils for Promotion or Vocational Guidance, South Pacific Commission, Goroka, PNG.

Mumford, M., & Owens, W. (1987). Methodology review: Principles, procedures, and findings in the application of background data measures. *Applied Psychological Measurement, 11*, 1-31.

Ord, 1. (1957). *Report on psychological investigations and findings concerning the assessment of natives of Papua-New Guinea during the period 15 June-13 Sept.,* 1957. Unpublished report, Headquarters, Australian Army, Northern Command, Victoria Barracks, Brisbane.

Ord, I. (1959). *The development of a test of cognitive capacity for indigenes of Papua and New Guinea.* Unpublished master's thesis, University of Queensland.

Ord, 1. (1966). Psychological services in Papua New Guinea. *Australian Psychologist, 1*, 100-105.

Ord, 1. (1967a). The New Guinea Performance Scale and its educational uses. *Papua New Guinea Journal of Education, 5,* 7-16.

Ord, 1. (1967b). *Manual for "New Guinea English Attainments and Skills" test series.* Unpublished report, 19 Psychology Unit, Murray Barracks, PNG Command, Port Moresby, PNG.

Ord, I. (1968). The P.I.R. Test and its derivatives. *Australian Psychologist, 2*, 137146.

Ord, I. (1971a). *Psychological test programmes and school certificate and university preliminary year results: Some findings in Papua and New Guinea.* (Research Report No. 2), Port Moresby: Psychological Services Branch, Public Service Board, PNG.

Ord, I. (197lb). Assessing the cognitive capacities of non-literate New Guinean adults. *New Guinea Psychologist Monograph Supplement*, *3*.

Preston, L., St George, A., & St George, R. (1974a). The origins and development of psychological tests in Papua New Guinea: A guide to name changes and test modifications. *New Guinea Psychologist*, 6, 96-103.

Preston, L., St George, A., & St George, R. (1974b, January). *Test information report no. 7: Pacific Reading Comprehension Test.* Port Moresby: Psychological Services Branch, Public Service Commission, PNG. (National Archives & Public Records Service of Papua New Guinea, Series 1050, Box 4253, File TIR.7)

Preston, L., St George, A., & St George, R. (1974c, January). *Test information report no. 8: Pacific Word Knowledge Test.* Port Moresby: Psychological Services Branch, Public Service Commission, PNG. (National Archives & Public Records Service of Papua New Guinea, Series 1050, Box 4253, File TIR8)

Preston, L., St George, A., & St George, R. (1974d, January). *Test information report no. 1: Pacific Reasoning Series Test.* Port Moresby: Psychological Services Branch, Public Service Commission, PNG. (National Archives & Public Records Service of Papua New Guinea, Series 1050, Box 4253, File TIRI)

Price, J. (1973). Reading comprehension among preliminary year students at UPNG. *New Guinea Psychologist, 5,* 106-109.

Price, J. (1984). A look at psychology in Papua New Guinea since independence. *South Pacific Journal of Psychology*, *1*, 1-15.

Psychological Services Branch (1982, February). *Manual for the school leaver testing programme.* Port Moresby: Public Service Commission, Papua New Guinea. (National Archives & Public Record Services of Papua New Guinea, Series 1044, Box 6964)

Psychological Services Branch (no date). *The validity of the 1981 Adcol long course selections.* Port Moresby: Public Service Commission, Papua New Guinea. (National Archives & Public Records Services of Papua New Guinea, Series 1044, Box 6964)

Sacket, P. (1987). Assessment centers and content validity: Some neglected issues. *Personnel Psychology, 40*, 13-25.

Schmitt, N., & Robertson, 1. (1990). Personnel selection. *Annual Review of Psychology, 41*, 289-319.

St George, R., & Preston, R. (1980). Psychological testing in Papua New Guinea: A critical appraisal of the work of the Psychological Services Branch. *Australian Psychologist, 15*, 57-71.

St George, R., & Preston, R. (1981a). Psychological testing in Papua New Guinea: A rejoinder to Hicks. *Australian Psychologist, 16,* 278-280.

S t George, R., & Preston, R. (1981b). Psychological testing in Papua New Guinea: A rejoinder to Hutton. *Australian Psychologist, 16*, 282-283.

Sulsky, L., & Balzer, W. (1988). Meaning and measurement of performance rating accuracy: Some methodological and theoretical concerns. *Journal of Applied Psychology, 73*, 497-506.

Wiesner, W., & Cronshaw, S. (1988). A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview. *Journal of Occupational Psychology*, 61, 275-290.